

Diseño muestral para el levantamiento de información de jóvenes colombianos en territorios del posconflicto

Diseño de muestreo para un levantamiento de información primaria a través de una encuesta en hogares a jóvenes de entre 15 y 24 años de edad en territorios del posconflicto en Colombia. La información recolectada permitirá conocer la percepción de los jóvenes en temas de desarrollo claves en el posconflicto así como sus expectativas y aspiraciones.

Este capítulo contiene los lineamientos metodológicos para la selección de una muestra probabilística de jóvenes, representativa del universo de jóvenes de los 170 municipios priorizados en el artículo 3 del decreto 893 de Mayo de 2017. Se consideran como variables de estratificación dos indicadores por municipio: (1) indicador de exposición a la violencia e (2) indicador del nivel de ruralidad. Se han seleccionado aleatoriamente 50 municipios en donde se levantará información de 2.300 jóvenes (en 2.300 hogares), residentes en la cabecera y en resto municipal de los municipios seleccionados. A continuación se describen los detalles del diseño muestral.

Universo en estudio

El universo en estudio es el conjunto de jóvenes entre 15 y 24 años de edad, residentes en viviendas particulares de los 170 municipios priorizados en el artículo 3 del decreto 893 de Mayo de 2017. Se excluye del estudio a jóvenes que viven en hogares colectivos o en la calle.

Población Estimada

Unidad de Observación. Joven definido en el universo en estudio.

El DANE ha proyectado al año 2017 un total de 1,34 millones de jóvenes en los 170 municipios priorizados. La tabla No 1 presenta la cantidad de jóvenes clasificados de acuerdo al nivel de exposición a la violencia (cuatro niveles) del municipio donde viven y de acuerdo al nivel de ruralidad¹ establecido por la Misión para la Transformación del Campo (dos niveles).

Tabla No 1. Cantidad de jóvenes por nivel de exposición a la violencia

Nivel de exposición a la violencia	Alta ruralidad	Baja ruralidad	Total	% Total
1 (bajo)	119.703	407.916	527.619	39%
2	134.600	229.300	363.900	27%
3	147.977	75.887	223.864	17%
4 (alto)	186.355	41.037	227.392	17%
Total	588.635	754.140	1.342.775	100%

Fuente: elaboración propia a partir de las proyecciones de población a 2017 del DANE

Para la construcción de los cuartiles de nivel de exposición a la violencia, se utilizaron los datos sobre casos de violencia en cabeza de la guerrilla y los paramilitares proporcionados por el Centro de Memoria Histórica. Se sumaron el total de casos para el periodo 2005-2017 y se dividieron sobre la población promedio en cada municipio entre 2005-2017 provista por el DANE. Finalmente, se construyeron los cuartiles de exposición al conflicto con base en este indicador.

Del total de jóvenes estudiados, cerca de 550 mil (el 41%) son jóvenes que residen en hogares ubicados en el “resto” del municipio (fuera de la cabecera municipal); el restante 59% vive en la cabecera municipal.

¹ Estratificación realizada por la Misión para la Transformación del Campo. Alta ruralidad corresponde a los municipios clasificados en “rural” y “rural disperso” y Baja ruralidad incluye al resto de municipios. Ver documento en <https://colaboracion.dnp.gov.co/CDT/Estudios%20Economicos/2015ago6%20Documento%20de%20Ruralidad%20-%20DDRS-MTC.pdf>

El nivel de reporte inherente al diseño muestral propuesto es la presentación de indicadores de resultados para los 170 municipios y aquellos que dan cuenta de los resultados por niveles de exposición a la violencia. Sin embargo, se puede considerar el ejercicio de reportar indicadores separando los jóvenes que habitan las cabeceras municipales de los jóvenes que habitan en el resto del municipio, verificando la robustez estadística del proceso.

Marcos de Muestreo

El marco de muestreo está conformado por la lista de 170 municipios priorizados (decreto 893 de 2017) y el Marco Geo-estadístico Nacional² elaborado por el DANE. Este marco de muestreo contiene el material necesario para operacionalizar el diseño de muestreo probabilístico propuesto en este documento.

Método de muestreo

El diseño de muestreo propuesto es probabilístico, estratificado y multietápico con selección aleatoria simple de unidades de muestreo.

La estratificación en la primera etapa (selección de municipios) considera cuatro (4) niveles de exposición a la violencia y dos (2) niveles de ruralidad para un total de ocho (8) estratos estadísticos. Buenaventura, Pradera y Florida, municipios del Valle del Cauca, y Argelia (Cauca) se incluyeron forzosamente en la muestra. La segunda etapa considera la *zona* (cabecera y resto del municipio -definiciones DANE-) como variable de sub-estratificación; en la cabecera municipal se seleccionarán bloques cartográficos y en el resto del municipio se seleccionarán veredas (incluidos los caseríos llamados centros poblados). Se tomarán todos los hogares dentro del bloque cartográfico seleccionado con el objetivo de encuestar a UN (1) joven por hogar (en aquellos donde habitan jóvenes de la población objetivo). En las veredas se realizará

² <https://geoportal.dane.gov.co/v2/?page=elementoDescargaMGN>

un trabajo similar. La selección del joven dentro del hogar se realizará utilizando el método de Kish. A continuación el detalle de las etapas de muestreo:

Etapas I. Selección de municipios. En cada uno de los 8 estratos estadísticos se realizó una selección aleatoria simple de municipios. El resultado en tamaño de muestra por estrato es el siguiente:

Tabla No 2. Cantidad de municipios seleccionados en la muestra

Nivel de exposición a la violencia	Nivel de ruralidad	Municipios en el universo	Municipios en la muestra
1 (bajo)		43	13
	Alto	23	7
	Bajo	20	6
2		43	13
	Alto	31	9
	Bajo	12	4
3		42	12
	Alto	37	9
	Bajo	5	3
4 (alto)		42	12
	Alto	38	10
	Bajo	4	2
Total		170	50

Fuente: elaboración propia

Dado que el proyecto contempla la visita a 50 municipios, la asignación del tamaño de muestra por estrato se realizó proporcional al tamaño del universo del estrato, considerando que cada estrato debe contener al menos dos municipios de inclusión probabilística para permitir los cálculos de varianza. Si fuera necesario reemplazar algunos municipios por la imposibilidad de implementar el operativo de campo (por temas de seguridad del personal de campo u otro motivo de fuerza mayor), se seleccionarán otros municipios siguiendo el mismo mecanismo de muestreo utilizado para seleccionar los iniciales.

Etapla II. Selección de bloques cartográficos y veredas. Dentro de cada municipio seleccionado, se considerarán dos sub-estratos estadísticos, (1) cabecera municipal y (2) resto del municipio. En la cabecera municipal se seleccionará una muestra aleatoria simple de bloques cartográficos y en el resto del municipio se seleccionará una muestra aleatoria simple de veredas (incluyendo los caseríos denominados centros poblados).

Dentro de cada bloque cartográfico y dentro de cada vereda se realizará el recorrido (censo) por todos los hogares, indagando sobre la existencia de jóvenes de la población objetivo. En cada hogar en donde habiten dos o más jóvenes de la población objetivo, la selección del encuestado se realizará a través de la tabla de Kish. Esta selección aleatoria del joven dentro del hogar es garantía de que los grupos de edad y el sexo (hombre/mujer) quedarán bien representados (se harán evaluaciones periódicas durante el operativo de campo para verificar el cumplimiento de estos criterios de microrepresentatividad).

Tamaño de Muestra, márgenes de Error y Nivel de confianza

El tamaño de la muestra (unidades finales de muestreo) es de 2.300 jóvenes. La tabla No 3 presenta el tamaño de muestra por nivel de exposición a la violencia, separando la cantidad de jóvenes que se encuestarán en la cabecera municipal y la cantidad de jóvenes que se encuestarán en el resto del municipio.

Tabla No 3. *Tamaño de muestra (n) por nivel de exposición a la violencia*

Nivel de exposición a la violencia	n cabecera municipal	n resto municipal	Total tamaño de muestra (n)
1 (bajo)	410	196	606
2	394	219	613
3	325	211	536
4 (alto)	201	344	545
Total	1.330	970	2.300

Fuente: elaboración propia

El tamaño de muestra mínimo por nivel de exposición a la violencia es de 536 jóvenes, lo cual garantiza un error máximo de muestreo de 4,7%, asumiendo la estimación de una proporción cercana a 0.5 y un efecto de diseño de 1,2. El error de muestreo para los indicadores de resultados del agregado de los 170 municipios priorizados es de 2,3%. Los intervalos de confianza para los parámetros en estudio se construirán con el 95% de confianza.

De estas 2.300 encuestas, 1.600 (el 70%) se realizarán en municipios que la Misión para la Transformación del Campo a identificado como municipios con alto nivel de ruralidad. Esta distribución del tamaño muestral es consecuencia de la asignación del tamaño de muestra por municipio, proporcional al logaritmo natural de la cantidad de jóvenes que habitan el municipio. Se ha definido también que el tamaño de la muestra en el resto del municipio será el 41% del tamaño de la muestra total (el tamaño de muestra en las cabeceras municipales es de 59% del tamaño de muestra total), esto debido a que ese es el indicador de participación poblacional de los jóvenes que viven fuera de la cabecera municipal. El tamaño de muestra por municipio y por zona (cabecera / resto) dentro del municipio se entrega en el anexo 1.

La expresión matemática utilizada para el cálculo de tamaño de muestra y para relacionar este con el error de muestreo, considerando que el parámetro de interés es la proporción p (indicador de expectativa de vida positiva de los jóvenes), es:

$$n = \frac{Np(1-p)def}{N(ESrel)^2p^2 + p(1-p)def}$$

Donde n es el tamaño de muestra, p es el parámetro de interés en el estudio, N es el tamaño de la población, def es el efecto de diseño y $ESrel$ es el Error Estándar Relativo máximo admitido en el estudio.

Método de selección de la muestra

Uno de los algoritmos más utilizados para seleccionar una muestra aleatoria simple de unidades estadísticas es el denominado “coordinado negativo”. Los siguientes son los pasos:

- (a) A cada unidad muestral asigne un número aleatorio entre “0” y “1”
- (b) Fije los valores aleatorios generados
- (c) Ordene ascendentemente las unidades de acuerdo al número generado dentro de cada estrato.
- (d) En cada estrato tome las primeras “ n ” unidades, donde “ n ” es el tamaño de muestra asignado a ese estrato.

Técnica de recolección

La labor de recolección de información en hogares se realizará a través de encuestas presenciales.

Anexo 1. Tamaño de muestra por municipio y zona del municipio

Código	Departamento	Municipio	Nivel (expo. violencia)	Nivel de ruralidad - MTC	Tamaño de muestra (n)	n cabecera	n resto
5665	Antioquia	SAN PEDRO DE URABA	1	Bajo	47	28	19
13654	Bolívar	SAN JACINTO	1	Alto	45	45	0
13212	Bolívar	CORDOBA	1	Alto	42	14	28
13442	Bolívar	MARIA LA BAJA	1	Bajo	49	29	20
18029	Caquetá	ALBANIA	1	Alto	39	20	19
19780	Cauca	SUAREZ	1	Alto	45	10	35
20570	Cesar	PUEBLO BELLO	1	Alto	46	15	31
20621	Cesar	LA PAZ	1	Alto	45	45	0
27450	Chocó	MEDIO SAN JUAN	1	Alto	44	21	23
47053	Magdalena	ARACATACA	1	Bajo	49	49	0
47001	Magdalena	SANTA MARTA	1	Bajo	62	62	0
70523	Sucre	PALMITO	1	Bajo	43	22	21
76563	Valle Del Cauca	PRADERA	1	Bajo	50	50	0
5604	Antioquia	REMEDIOS	2	Alto	47	22	25
5120	Antioquia	CACERES	2	Alto	48	14	34
5250	Antioquia	EL BAGRE	2	Bajo	50	34	16
13042	Bolívar	ARENAL	2	Alto	45	16	29
13744	Bolívar	SIMITI	2	Alto	45	30	15
19318	Cauca	GUAPI	2	Alto	48	38	10
19050	Cauca	ARGELIA	2	Alto	46	9	37
27800	Chocó	UNGUIA	2	Alto	43	18	25
23466	Córdoba	MONTELIBANO	2	Bajo	52	52	0
52490	Nariño	OLAYA HERRERA	2	Alto	48	20	28
52390	Nariño	LA TOLA	2	Alto	43	43	0
70230	Sucre	CHALAN	2	Bajo	37	37	0
76109	Valle Del Cauca	BUENAVENTURA	2	Bajo	61	61	0
5234	Antioquia	DABEIBA	3	Alto	45	24	21
5790	Antioquia	TARAZA	3	Bajo	49	49	0
18860	Caquetá	VALPARAISO	3	Alto	42	18	24
18479	Caquetá	MORELIA	3	Alto	36	24	12
19455	Cauca	MIRANDA	3	Bajo	49	49	0
27150	Chocó	CARMEN DEL DARIEN	3	Alto	38	12	26
50450	Meta	PUERTO CONCORDIA	3	Alto	45	29	16

Código	Departamento	Municipio	Nivel (expo. violencia)	Nivel de ruralidad - MTC	Tamaño de muestra (n)	n cabecera	n resto
52250	Nariño	EL CHARCO	3	Alto	49	17	32
52405	Nariño	LEIVA	3	Alto	42	18	24
52418	Nariño	LOS ANDES	3	Alto	45	23	22
73616	Tolima	RIOBLANCO	3	Alto	46	12	34
76275	Valle Del Cauca	FLORIDA	3	Bajo	50	50	0
81794	Arauca	TAME	4	Alto	50	25	25
81736	Arauca	SARAVENA	4	Bajo	50	40	10
18410	Caquetá	LA MONTAÑITA	4	Alto	46	13	33
19821	Cauca	TORIBIO	4	Alto	47	4	43
19212	Cauca	CORINTO	4	Bajo	48	25	23
50350	Meta	LA MACARENA	4	Alto	47	8	39
50577	Meta	PUERTO LLERAS	4	Alto	40	16	24
52612	Nariño	RICAURTE	4	Alto	45	8	37
52540	Nariño	POLICARPA	4	Alto	44	10	34
54250	Norte De Santander	EL TARRA	4	Alto	42	23	19
54670	Norte De Santander	SAN CALIXTO	4	Alto	43	9	34
86569	Putumayo	PUERTO CAICEDO	4	Alto	43	20	23
Total					2.300	1.330	970